

**Nucleic acid molecule comprising a
nucleic acid sequence which codes for a haemocyanin,
and comprising at least one intron sequence**

The present invention relates to a nucleic acid molecule comprising a nucleic acid sequence which codes for a haemocyanin, a haemocyanin domain or a fragment with the immunological properties of at least one domain of haemocyanin, and comprising at least one intron sequence, constructs which comprise such molecules, host cells which comprise the nucleic acid sequences or the constructs, processes for the preparation of haemocyanin polypeptides, and recombinant haemocyanin polypeptides.

Haemocyanin is a blue copper protein which occurs in a freely dissolved form in the blood of numerous molluscs and arthropods and transports oxygen. Of the molluscs, the cephalopods, chitons, most gastropods and some bivalves contain haemocyanin. Among the arthropods, haemocyanin is typical of arachnids, xiphosurans, malacostracan crustaceans and *Scutigera*. Numerous species of insects contain proteins which are derived from haemocyanin. Haemocyanins are present in the extracellular medium and float in the haemolymph.

While arthropod haemocyanin has a maximum diameter of 25 nm under an electron microscope and a subunit has a molecular weight of 75,000 Da, mollusc cyanins are much larger. Thus e.g. the haemocyanin of *Megathura* has a diameter of 35 nm and is composed of 2 subunits. Each subunit has a molecular weight of approx. 400,000 Da and is divided into eight oxygen-binding domains, each of which has a molecular weight of approx. 50,000. The domains differ immunologically. These domains can be liberated from the subunit by limited proteolysis.

The haemocyanin of gastropods visible under an electron microscope has a molecular weight of approx. 8 million Da and is a di-decamer. In contrast to this, the haemocyanin of cephalopods is arranged as an isolated decamer, which also differs significantly from the haemocyanin of gastropods in the quaternary structure.

The haemocyanin of the Californian keyhole limpet *Megathura crenulata* is of particular immunological interest. The haemocyanin is therefore also called keyhole limpet haemocyanin (KLH). Haemocyanins are very potent antigens. Immunization of a vertebrate leads to a non-specific activation of the immune system which to date is not very well understood. By the general activation of the immune system, it is then possible also to achieve an immune reaction to other foreign structures which have previously been tolerated. KLH is used above all as a hapten carrier in order thus to achieve the formation of antibodies against the hapten.

In addition to *Megathura crenulata*, the abalone *Haliotis tuberculata* also belongs to the Archaeogastropoda group, which is relatively old in respect of evolution. It is known that *Haliotis* also produces haemocyanin.

KLH is a mixture of two different haemocyanins, which are called KLH1 and KLH2. The subunit of KLH1 is a 390 kDa polypeptide which consists of eight globular domains called 1 a to 1 h according to their sequence in the subunit. On the other hand, KLH2 has a molecular weight of 350 kDa and according to the most recent data also contains 8 domains, called 2 a to 2 h. *In vivo* every type of subunit forms homo-oligomers, while no hetero-oligomers have been observed.

Amino-terminal, internal and carboxy-terminal domains have been obtained by limited proteolysis and crossed immunoelectrophoresis of the subunit of KLH1 and KLH2, and their amino-terminal sequences has been determined (Söhngen et al., Eur. J. Biochem. 248 (1997), 602-614; Gebauer et al., Zoology 98(1994), 51-68). However, the resulting sequences do not allow designing of sequence-specific primers and/or probes which promise success for hybridization with genomic DNA. Although both KLH types have been known since 1991 and 1994 respectively, it has so far not been possible to clarify the primary structure.

At the DNA level, in respect of molluscs only the cDNA sequence of the haemocyanin subunit from the cephalopod *Octopus dofleini* is so far known (Miller et al., J. Mol. Biol. 278 (1998), 827-842). *Octopus dofleini* is phylogenetically very far removed from the

archaeogastropods. A haemocyanin gene sequence from molluscs is so far not known at all.

As described by Miller et al. *supra*, it is difficult both to isolate a single functional domain (functional unit = domain; also called functional domain) and to obtain tissue which is suitable for purification of mRNA for cDNA sequencing.

There is a further difficulty in the analysis of the haemocyanin from *Megathura crenulata* in that the test animals must have reached an age of 4 to 8 years for haemolymph to be taken from them in the first place. After the haemolymph has been taken, haemocyanin is not subsequently produced in these animals. It is not yet known how haemocyanin synthesis could be stimulated. Furthermore, culture of *Megathura* is extremely expensive, since special flow basins are required for this.

It is therefore an object of the present invention to provide means and ways in order to be able to produce haemocyanin and/or domains thereof in a sufficient amount and inexpensively. This includes the further object of providing a process with which this haemocyanin can be prepared.

This object is achieved according to the invention by a nucleic acid molecule comprising a nucleic acid sequence which codes for a haemocyanin, a haemocyanin domain or a functional fragment thereof with the immunological properties of at least one domain of a haemocyanin, and comprising at least one intron sequence, the nucleic acid sequence being selected from

- (a) nucleic acid sequences which are selected from the group consisting of the DNA sequences shown below or the corresponding RNA sequences or which contain these:

SEQ ID NO:1 (HtH1 domain a + signal peptide),

SEQ ID NO:2 (HtH1 domain b),

SEQ ID NO:3 (HtH1 domain c),

SEQ ID NO:4 (HtH1 domain d),

SEQ ID NO:5 (HtH1 domain e),
SEQ ID NO:6 (HtH1 domain f),
SEQ ID NO:7 (HtH1 domain g),
SEQ ID NO: 8 (HtH1 domain h),
SEQ ID NO:9 (partial HtH2 domain b),
SEQ ID NO:10 (HtH2 domain c),
SEQ ID NO:11 (HtH2 domain d),
SEQ ID NO:12 (HtH2 domain e),
SEQ ID NO:13 (HtH2 domain f),
SEQ ID NO:14 (HtH2 domain g),
SEQ ID NO:15 (HtH2 domain h),
SEQ ID NO:16 (partial KLH1 domain b),
SEQ ID NO:17 (KLH1 domain c),
SEQ ID NO:18 (KLH1 domain d),
SEQ ID NO:19 (partial KLH1 domain e),
SEQ ID NO:20 (KLH2 domain b),
SEQ ID NO:21 (KLH2 domain c),
SEQ ID NO:22 (partial KLH2 domain d),
SEQ ID NO:23 (KLH2 domain g),
SEQ ID NO:24 (partial KLH2 domain h),
SEQ ID NO:49 (HtH1 domain a' + signal peptide),
SEQ ID NO:50 (partial HtH2 domain a),
SEQ ID NO:51 (HtH2 domain b'),
SEQ ID NO:52 (HtH2 domain d'),
SEQ ID NO:53 (HtH2 domain e'),
SEQ ID NO:54 (KLH1 domain e'),
SEQ ID NO:55 (KLH1 domain f),
SEQ ID NO:56 (KLH1 domain g),
SEQ ID NO:57 (KLH2 domain b'),
SEQ ID NO:58 (KLH2 domain c'),
SEQ ID NO:59 (KLH2 domain d'),
SEQ ID NO:60 (KLH1 domain e),
SEQ ID NO:61 (KLH2 domain f),

SEQ ID NO:62 (KLH2 domain g'),
SEQ ID NO:80 (HtH1 domain a" + signal peptide),
SEQ ID NO:81 (HtH1 domain b"),
SEQ ID NO:82 (HtH1 domain c"),
SEQ ID NO:83 (HtH1 domain d"),
SEQ ID NO:84 (HtH1 domain e"),
SEQ ID NO:85 (HtH1 domain f"),
SEQ ID NO:86 (HtH1 domain g"),
SEQ ID NO:87 (HtH1 domain h"),
SEQ ID NO:88 (partial HtH2 domain a"),
SEQ ID NO:89 (HtH2 domain b"),
SEQ ID NO:90 (HtH2 domain c"),
SEQ ID NO:91 (HtH2 domain d"),
SEQ ID NO:92 (HtH2 domain e"),
SEQ ID NO:93 (HtH2 domain f"),
SEQ ID NO:94 (HtH2 domain g"),
SEQ ID NO:95 (HtH2 domain h"),
SEQ ID NO:96 (partial KLH1 domain b"),
SEQ ID NO:97 (KLH1 domain c"),
SEQ ID NO:98 (KLH1 domain d"),
SEQ ID NO:99 (KLH1 domain e"),
SEQ ID NO:100 (KLH1 domain f"),
SEQ ID NO:101 (KLH1 domain g"),
SEQ ID NO:102 (KLH2 domain b"),
SEQ ID NO:103 (KLH2 domain c"),
SEQ ID NO:104 (KLH2 domain d"),
SEQ ID NO:105 (KLH2 domain e"),
SEQ ID NO:106 (KLH2 domain f"),
SEQ ID NO:107 (KLH2 domain g"),
SEQ ID NO:108 (partial KLH2 domain h"),
SEQ ID NO:157 (complete HtH2 domain a),

- (b) nucleic acid sequences which hybridize with the counter-strand of a nucleic acid sequence according to (a) and code for a polypeptide which has the immunological properties of at least one domain of a haemocyanin;
- (c) nucleic acid sequences which on the basis of the genetic code are degenerated to the DNA sequences defined under (a) and (b) and code for a polypeptide which has the immunological properties of at least one domain of a haemocyanin;
- (d) nucleic acid sequences which hybridize with one of the nucleic acid sequences described under (a) to (c) and the counter-strand of which codes for a polypeptide which has the immunological properties of at least one domain of a haemocyanin;
- (e) nucleic acid sequences which are at least 60% homologous to one of the nucleic acid sequences described under (a);
- (f) variants of the sequences described under (a) to (e), the variants containing additions, deletions, insertions or inversions and coding for a polypeptide which has the immunological properties of at least one domain of haemocyanin; and
- (g) combinations of several of the DNA sequences described under (a) to (f).

Preferably, the intron sequence is selected from the following sequences:

- (i) nucleic acid sequences which are selected from the group consisting of the DNA sequences shown below or the corresponding RNA sequences or which contain these:

SEQ ID NO:109 (HtH1 intron 1S-1/1S-2),
SEQ ID NO:110 (HtH1 intron 1S-2/1A-1),
SEQ ID NO:111 (HtH1 intron 1A-1/1A-2),
SEQ ID NO:112 (HtH1 intron 1A-2/1A-3),
SEQ ID NO:113 (HtH1 intron 1A-3/1A-4),
SEQ ID NO:114 (HtH1 intron 1A-4/1B),

SEQ ID NO:115 (HtH1 intron 1B/1C),
SEQ ID NO:116 (HtH1 intron 1C/1D),
SEQ ID NO:117 (HtH1 intron 1D/1E),
SEQ ID NO:118 (HtH1 intron 1E/1F-1),
SEQ ID NO:119 (HtH1 intron 1F-1/1F-2),
SEQ ID NO:120 (HtH1 intron 1F-2/1G-1),
SEQ ID NO:121 (HtH1 intron 1F-1/1G-2),
SEQ ID NO:122 (HtH1 intron 1G-2/1G-3),
SEQ ID NO:123 (HtH1 intron 1G-3/1H),
SEQ ID NO:124 (intron in the 3'UTR of HtH1),
SEQ ID NO:125 (HtH2 intron 2A-1/2A-2),
SEQ ID NO:126 (HtH2 intron 2A-1/2A-3),
SEQ ID NO:127 (HtH2 intron 2A-1/2A-4),
SEQ ID NO:128 (HtH2 intron 2A-4/2B),
SEQ ID NO:129 (HtH2 intron 2B/2C),
SEQ ID NO:130 (HtH2 intron 2C/2D),
SEQ ID NO:131 (HtH2 intron 2D/2E),
SEQ ID NO:132 (HtH2 intron 2E/2F-1),
SEQ ID NO:133 (HtH2 intron 2F-1/2F-2),
SEQ ID NO:134 (HtH2 intron 2F-2/2GF-1),
SEQ ID NO:135 (HtH2 intron 2G-1/2G-2),
SEQ ID NO:136 (HtH2 intron 2G-2/2G-3),
SEQ ID NO:137 (HtH2 intron 2G-3/2H),
SEQ ID NO:138 (intron in the 3'UTR of HtH2),
SEQ ID NO:139 (KLH1 intron 1B/1C),
SEQ ID NO:140 (KLH1 intron 1C/1D),
SEQ ID NO:141 (KLH1 intron 1D/1E),
SEQ ID NO:142 (KLH1 intron 1E/1F),
SEQ ID NO:143 (KLH1 intron 1F-1/1F-2),
SEQ ID NO:144 (KLH1 intron 1F-2/1G-1),
SEQ ID NO:145 (KLH1 intron 1G-1/1G-2),
SEQ ID NO:146 (KLH1 intron 1G-2/1G-3),
SEQ ID NO:147 (KLH2 intron 2B/2C),

SEQ ID NO:148 (KLH2 intron 2C/2D),
SEQ ID NO:149 (KLH2 intron 2D/2E),
SEQ ID NO:150 (KLH2 intron 2E/2F),
SEQ ID NO:151 (KLH2 intron 2F),
SEQ ID NO:152 (KLH2 intron 2F-2/2G),
SEQ ID NO:153 (KLH2 intron 2G-1/2G-2),
SEQ ID NO:154 (KLH2 intron 2G-2/2G-3),
SEQ ID NO:155 (KLH2 intron 2G/2H);

- (ii) nucleic acid sequences which hybridize with the counter-strand of a nucleic acid sequence according to (i);
- (iii) nucleic acid sequences which are at least 60% homologous to one of the nucleic acid sequences described under (i);
- (iv) variants of the sequences described under (i) to (iii), the variants containing additions, deletions, insertions or inversions with respect to the sequences described under (i) to (iii); and
- (v) combinations of several of the DNA sequences described under (i) to (iv).

Some terms are explained in more detail below in order to clarify how they are to be understood in connection with the present application.

The term "haemocyanin" as used below in the description includes complete haemocyanin, haemocyanin domains and/or fragments, haemocyanin mutants and fusion proteins. In respect of fusion proteins, these include, in particular, those in which the fusion comprises haemocyanin and antigens.

"Domains" are understood as meaning functional partial sequences of the haemocyanin subunits which can be separated from one another, for example, by limited proteolysis. They can furthermore have different immunological properties.

The "immunological properties of at least one domain of haemocyanin" means the property of a polypeptide of inducing, in the same manner as at least one domain of haemocyanin, an immunological response of the recipient immunized with the polypeptide. "Immunological response" here is understood as meaning T and/or B cell responses to haemocyanin epitopes, such as, for example, an antibody production. The immunological reaction can be observed, for example, by immunization of a mammal, such as e.g. a mouse, a rat or a rabbit, with the corresponding polypeptide and comparison of the immune response to the polypeptide used for the immunization with the immune response to natural haemocyanins.

The term "intron sequence" refers either to a sequence interrupting an eukaryotic gene or to the corresponding sequence in the RNA transcript. The intron sequence(s) and the coding sequence(s) are transcribed together; the intron transcript or transcripts are then deleted to obtain the functional RNA.

According to the invention, the term "antigen" includes both haptens and weak and potent antigens. Haptens are characterized in that they are substances of low molecular weight (less than 4,000 Da), but without being coupled to a carrier molecule are not capable of inducing an immunological reaction. Weak antigens are substances which can themselves already induce an immunological reaction and of which the potential to be able to induce an immunological reaction can be increased further by coupling with a carrier molecule at the protein and/or DNA level.

"His tag" means a sequence of at least 6 histidine amino acids which, by corresponding cloning and fusion with an expressible sequence, leads to a fusion protein which has at least 6 His residues on the NH₂ terminus and can easily be purified by complexing with an Ni²⁺ column.

"Cloning" is intended to include all cloning methods known in the prior art which could be employed here but which are not all described in detail because they belong to the obvious hand tools of the skilled person.

"Variants" of a nucleic acid sequences include additions, deletions, insertions or inversions and code for a polypeptide which has the immunological properties of at least one domain of a haemocyanin. Variants can be synthetic or natural. Allelic variants are an example of natural variants.

"Recombinant expression in a suitable host cell" is to be understood as meaning all the expression methods known in the prior art in known expression systems which could be employed here but which are not all described in detail because they belong to the obvious hand tools of the skilled person.

The nucleic acid sequence contained in the nucleic acid molecule according to the invention can be genomic DNA, cDNA or synthetic DNA, synthetic DNA sequences also being understood as meaning those which comprise modified internucleoside bonds.

The nucleic acid sequences can furthermore be RNA sequences, which may be necessary e.g. for expression by means of recombinant vector systems. The nucleic acid sequences according to (b) are obtainable, for example, by using a detectably marked probe which corresponds to one of the sequences described under (a) or a fragment, or a counter-strand thereof for screening cDNA/genomic DNA libraries from molluscs or arthropods. The mRNA on which the cDNA library is based is preferably to be obtained from mollusc tissues which express haemocyanin to a particularly high degree, such as e.g. mantle tissue from gastropods and branchial gland tissue from cephalopods.

Positive genomic DNA clones are identified by standard methods. Cf. Maniatis et al., Molecular Cloning (1989) Cold Spring Harbor Laboratory Press.

In a preferred embodiment, the hybridization described under (b), (d) or (ii) is carried out under stringent conditions. Stringent hybridization conditions are e.g. 68°C overnight in 0.5 x SSC; 1% blocking reagent (Boehringer Mannheim); 0.1% sodium lauryl sarcosinate and subsequent washing with 2 x SSC; 0.1% SDS.

In a preferred embodiment, nucleic acid molecules comprising a nucleic acid sequence which is at least 60% homologous to one of the nucleic acid sequences described under

(a) are provided. Nucleic acid sequences which are at least 80% homologous to one of the nucleic acid sequences described under (a) are preferred. Nucleic acid sequences which are at least 90% homologous to one of the nucleic acid sequences described under (a) are particularly preferred. In particular, the nucleic acid sequences are at least 95% homologous to one of the nucleic acid sequences described under (a).

In a further preferred embodiment, nucleic acid sequences comprising at least one intron sequence which is at least 60% homologous to one of the nucleic acid sequences described under (i) are provided. Intron sequence(s) which are at least 80% homologous to one of the nucleic acid sequences described under (i) are preferred. Intron sequence(s) which are at least 90% homologous to one of the nucleic acid sequences described under (i) are particularly preferred. In particular, the intron sequence(s) are at least 95% to one of the nucleic acid sequences described under (i).

According to the invention, the term "homology" means homology at the DNA level, which can be determined by known methods, e.g. computer-assisted sequence comparisons (Basic local alignment search tool, S.F. Altschul et al., *J. Mol. Biol.* 215 (1990), 403-410).

The term "homology" known to the skilled person describes the degree to which two or more nucleic acid molecules are related, this being determined by the concordance between the sequences. The percentage of "homology" is obtained from the percentage of identical regions in two or more sequences, taking into account gaps or other sequence peculiarities.

The homology of nucleic acid molecules which are related to one another can be determined with the aid of known methods. As a rule, special computer programs with algorithms which take account of the particular requirements are employed.

Preferred methods for the determination of homology initially produce the greatest concordance between the sequences analysed. Computer programs for determination of the homology between two sequences include, but are not limited to, the GCG program package, including GAP (Devereux, J., et al., *Nucleic Acids Research* 12 (12): 387 (1984); Genetics Computer Group University of Wisconsin, Madison, (WI)); BLASTP, BLASTN and FASTA

(Altschul, S. et al., *J. Mol. Biol.* 215:403-410 (1990)). The BLASTX program can be obtained from the National Centre for Biotechnology Information (NCBI) and from other sources (BLAST Handbook, Altschul S., et al., NCB NLM NIH Bethesda MD 20894; Altschul, S., et al., *J. Mol. Biol.* 215:403-410 (1990)). The known Smith Waterman algorithm can also be used for determining homologies.

Preferred parameters for the comparison of nucleic acid sequences include the following:

Algorithm:	Needleman and Wunsch, <i>J. Mol. Biol.</i> 48:443-453 (1970)
Comparison matrix:	Concordance (matches) = + 10
	Non-concordance (mismatch) = 0
Gap penalty:	50
Gap length penalty:	3

The GAP program is also suitable for use with the above parameters. The above parameters are the default parameters for nucleic acid sequence comparisons.

Further algorithms, gap opening penalties, gap extension penalties and comparison matrices by way of example, including those mentioned in the Program Handbook, Wisconsin Package, version 9, September 1997, can be used. The choice depends on the comparison to be made and furthermore on whether the comparison is to be made between sequence pairs, in which case GAP or Best Fit are preferred, or between a sequence and a comprehensive sequence databank, in which case FASTA or BLAST are preferred.

A concordance of 60% determined with the abovementioned algorithm is designated 60% homology in the context of this application. The same applies accordingly to higher degrees of homology.

In a preferred embodiment, the DNA sequence according to the invention is a combination of several of the DNA sequences described under (a) to (f) and at least one intron sequence, and can be obtained by fusion and optionally cloning, which are known to the skilled person. These combinations are of particular interest, since the encoded polypeptides are particularly immunogenic. Combinations which contain several or all of

the domains in the sequence (a to h) which occurs naturally in the subunit are particularly preferred. Embodiments in which, after deletion of the intron sequence(s), the nucleic acid sequences which code for the domains are coupled to one another directly in frame are particularly preferred.

Constructs which comprise the nucleic acid molecules according to the invention are furthermore provided. In a preferred embodiment, the construct according to the invention comprises a promoter which is suitable for expression, the nucleic acid sequence being under the control of the promoter. The choice of promoter depends on the expression system used for expression. Generally, constitutive promoters are preferred, but inducible promoters, such as e.g. the metallothionein promoter, are also possible.

In a further preferred embodiment, the construct furthermore comprises an antigen-coding nucleic acid sequence which is bonded directly to the haemocyanin nucleic acid according to the invention. The antigen-coding sequence can be located both 5' and 3' relative to the haemocyanin sequence or also on both ends. It either follows the haemocyanin sequence directly in the same reading frame, or is coupled to it by a nucleic acid linker, the reading frame being preserved. By fusion of the antigen-coding sequence with the haemocyanin sequence the formation of a fusion protein in which the antigen-coding sequence is bonded covalently to the haemocyanin sequence is intended. The antigen according to the invention is a medically relevant antigen, which is selected, for example, from: tumour antigens, virus antigens and antigens of bacterial or parasitic pathogens. Tumour antigens can be, for example, Rb and p53. The virus antigens preferably originate from immunologically relevant viruses, such as e.g. influenza virus, hepatitis virus and HIV. Pathogen antigens are, inter alia, those from mammalian pathogens, in particular organisms which are pathogenic to humans, such as e.g. *Plasmodium*. Bacterial antigens can originate e.g. from *Klebsiella*, *Pseudomonas*, *E. coli*, *Vibrio cholerae*, *Chlamydia*, *Streptococci* or *Staphylococci*.

In another preferred embodiment, the construct furthermore comprises at least a part of a vector, in particular regulatory regions, the vector being selected from: bacteriophages, such as λ derivatives, adenoviruses, vaccinia viruses,

baculoviruses, SV40 viruses and retroviruses, preferably MoMuLV (Moloney murine leukaemia virus).

A construct which additionally comprises a His tag-coding DNA sequence, which, when expressed, leads to the formation of a fusion protein with a His tag on the NH₂ terminus of the haemocyanin, facilitating purification of the protein on a nickel column by chelate formation, is furthermore preferred.

The invention furthermore provides host cells which contain the construct and which are suitable for expression of the construct. Numerous prokaryotic and eukaryotic expression systems are known in the prior art, the host cells being selected, for example, from prokaryotic cells, such as *E. coli* or *B. subtilis*, from eukaryotic cells, such as yeast cells, plant cells, insect cells and mammalian cells, e.g. CHO cells, COS cells or HeLa cells, and derivatives thereof. For example certain CHO production lines of which the glycosylation patterns are altered compared with CHO cells are known in the prior art. The haemocyanins obtained using glycosylation-deficient or glycosylation-reduced host cells possibly have additional epitopes which are otherwise not accessible to the immune system of the recipient in the case of complete glycosylation, so that haemocyanins with a reduced glycosylation under certain circumstances have an increased immunogenicity. From plant cells transformed with the construct according to the invention it is possible to produce transgenic plants or plant cell cultures which produce haemocyanin polypeptides, for example tobacco, potato, tomato, sugar beet, soya bean, coffee, pea, bean, rape, cotton, rice or maize plants or plant cell cultures.

The present invention also relates to a process for the preparation of a haemocyanin polypeptide. For this, the nucleic acid molecule according to the invention and/or the construct is expressed in a suitable host cell and the protein is isolated from the host cell or the medium by means of conventional processes.

Numerous processes for expression of DNA sequences are known to the skilled person; compare Recombinant Gene Expression Protocols in Methods in Molecular Biology, volume 62, Humana Press Totowa New Jersey (1995). The expression can be both constitutive and inducible, inducers such as, for example, IPTG and Zn²⁺ being known to

the skilled person. If a His tag has been fused on to the NH₂ terminus of the haemocyanin, the haemocyanin prepared can be purified by chelate formation on a nickel column. Processes for the purification of haemocyanin, in particular KLH, are to be found in Harris et al., *Micron* 26 (1995), 201-212. The haemocyanin is preferably purified by ion exchange chromatography and/or gel filtration chromatography. The procedure for these measures is known to the skilled person.

In another preferred embodiment, the haemocyanin prepared according to the invention is modified. The modifications include di-, oligo- and polymerization of the monomeric starting substance, for example by crosslinking, e.g. by means of dicyclohexylcarbodiimide or pegylation or association (self assembly). The di-, oligo- and polymers prepared in this way can be separated from one another by gel filtration. The formation of decamers, didecamers or multidecamers is intended in particular. Further modifications include side chain modifications, for example of ϵ -amino-lysine residues of the haemocyanin, or amino- or carboxy-terminal modifications. Modification of the haemocyanin by covalent bonding to an antigen is particularly preferred, it being possible for the antigen to be reacted stoichiometrically or non-stoichiometrically with the haemocyanin. The antigen is preferably selected from tumour antigens, virus antigens and pathogen antigens, as mentioned above. Further modifications include post-translational events, e.g. glycosylation or partial or complete deglycosylation of the protein.

In a preferred embodiment, the haemocyanin obtained by recombinant expression in prokaryotes or glycosylation-deficient eukaryotes is non-glycosylated. Haemocyanin which is glycosylated by recombinant expression in eukaryotes which are capable of glycosylation, such as yeast cells, plant cells, insect cells or mammalian cells, such as CHO cells or HeLa cells, is also possible according to the invention.

Haemocyanin polypeptides which comprise an amino acid sequence, the amino acid sequence being coded by one or more of the nucleic acid molecules according to the invention, are provided in another embodiment,

Haemocyanin polypeptides which comprise at least one amino acid sequence selected from the following group:

SEQ ID NO:25 (HtH1 domain a + signal peptide),
SEQ ID NO:26 (HtH1 domain b),
SEQ ID NO:27 (HtH1 domain c),
SEQ ID NO:28 (HtH1 domain d),
SEQ ID NO:29 (HtH1 domain e),
SEQ ID NO:30 (HtH1 domain f),
SEQ ID NO:31 (HtH1 domain g),
SEQ ID NO:32 (HtH1 domain h),
SEQ ID NO:33 (partial HtH2 domain b),
SEQ ID NO:34 (HtH2 domain c),
SEQ ID NO:35 (HtH2 domain d),
SEQ ID NO:36 (HtH2 domain e),
SEQ ID NO:37 (HtH2 domain f),
SEQ ID NO:38 (HtH2 domain g),
SEQ ID NO:39 (HtH2 domain h),
SEQ ID NO:40 (partial KLH1 domain b),
SEQ ID NO:41 (KLH1 domain c),
SEQ ID NO:42 (partial KLH1 domain d),
SEQ ID NO:43 (partial KLH1 domain e),
SEQ ID NO:44 (KLH2 domain b),
SEQ ID NO:45 (KLH2 domain c),
SEQ ID NO:46 (partial KLH2 domain d),
SEQ ID NO:47 (KLH2 domain g),
SEQ ID NO:48 (partial KLH2 domain h),
SEQ ID NO:63 (HtH1 domain a' + signal peptide),
SEQ ID NO:64 (HtH1 domain h'),
SEQ ID NO:65 (partial HtH2 domain a),
SEQ ID NO:156 (complete HtH2 domain a),
SEQ ID NO:66 (HtH2 domain b'),
SEQ ID NO:67 (HtH2 domain d'),

SEQ ID NO:68 (HtH2 domain e'),
SEQ ID NO:69 (partial KLH1 domain b'),
SEQ ID NO:70 (KLH1 domain e'),
SEQ ID NO:71 (KLH1 domain f'),
SEQ ID NO:72 (KLH1 domain g'),
SEQ ID NO:73 (KLH1 domain h'),
SEQ ID NO:74 (KLH2 domain b'),
SEQ ID NO:75 (KLH2 domain c'),
SEQ ID NO:76 (KLH2 domain d'),
SEQ ID NO:77 (KLH2 domain e'),
SEQ ID NO:78 (KLH2 domain f'),
SEQ ID NO:79 (KLH2 domain g'),
SEQ ID NO:158 (partial KLH2 domain h'),

or a fragment of one of these sequences which has the immunological properties of at least one domain of haemocyanin are preferred.

The invention also includes haemocyanin polypeptides of which the sequence shows at least 60% or 70%, preferably at least 80%, particularly preferably at least 90% or 95% homology to one of the amino acid sequences according to SEQ ID NO:25 to 48 and SEQ ID NO:63 to 79 over a partial region of at least 90 amino acids.

In this connection, the expression "at least 70%, preferably at least 80%, particularly preferably at least 90% homology" relates to concordance at the amino acid sequence level, which can be determined by known methods, e.g. computer-assisted sequence comparisons (Basic local alignment search tool, S.F. Altschul et al., J. Mol. Biol. 215 (1990), 403-410).

The term "homology" known to the skilled person describes here the degree to which two or more polypeptide molecules are related, this being determined by the concordance between the sequences, concordance being understood as meaning both identical concordance and conservative amino acid exchange. The percentage of

"homology" is obtained from the percentage of regions in concordance in two or more sequences, taking into account gaps or other sequence peculiarities.

The expression "conservative amino acid exchange" relates to an exchange of an amino acid residue for another amino acid residue, where the exchange does not lead to a change in polarity or charge. An example of a conservative amino acid exchange is the exchange of a non-polar amino acid residue for another non-polar amino acid residue.

The homology of polypeptide molecules which are related to one another can be determined with the aid of known methods. As a rule, special computer programs with algorithms which take account of the particular requirements are employed. Preferred methods for the determination of homology initially produce the greatest concordance between the sequences analysed. Computer programs for determination of the homology between two sequences include, but are not limited to, the GCG program package, including GAP (Devereux, J., et al., *Nucleic Acids Research* 12 (12): 387 (1984); Genetics Computer Group University of Wisconsin, Madison, (WI)); BLASTP, BLASTN and FASTA (Altschul, S. et al., *J. Molec. Biol* 215:403/410 (1990)). The BLAST X program can be obtained from the National Centre for Biotechnology Information (NCBI) and from other sources (BLAST Handbook, Altschul S., et al., NCB NLM NIH Bethesda MD 20894; Altschul, S., et al., *J. Mol.* 215:403/410 (1990)). The known Smith Waterman algorithm can also be used for determining homology.

Preferred parameters for the sequence comparison include the following:

Algorithm:	Needleman and Wunsch, <i>J. Mol. Biol</i> 48:443-453 (1970)
Comparison matrix:	BLOSUM 62 of Henikoff and Henikoff, <i>Proc. Natl. Acad. Sci. USA</i> 89:10915-10919 (1992)
Gap penalty:	12
Gap length penalty:	4
Similarity threshold:	0

The GAP program is also suitable for use with the above parameters. The above parameters are the standard parameters (default parameters) for amino acid sequence comparisons where gaps at the ends do not reduce the homology value. If sequences are very short compared with the reference sequence, it may furthermore be necessary to increase the expected value to up to 100,000 and where appropriate to reduce the word size down to 2.

Further algorithms, gap opening penalties, gap extension penalties and comparison matrices by way of example, including those mentioned in the Programm-Handbuch, Wisconsin-Paket [Program Handbook, Wisconsin Package], version 9, September 1997, can be used. The choice depends on the comparison to be made and furthermore on whether the comparison is to be made between sequence pairs, in which case GAP or best fit are preferred, or between a sequence and a comprehensive sequence database, in which case FASTA or BLAST are preferred.

A concordance of 60% determined with the above mentioned algorithm is designated 60% homology in the context of this Application. The same applies accordingly to higher degrees of homology.

In another embodiment, the invention provides haemocyanin polypeptides which are obtainable by the recombinant preparation method or modifications thereof.

Preferred haemocyanin polypeptides are those which comprise each of the sequences SEQ ID NO: 25 to 32, it being possible for the sequence with SEQ ID NO:25 to be replaced by SEQ ID NO:63 and/or SEQ ID NO:32 to be replaced by SEQ ID NO:64. Haemocyanin polypeptides which are also preferred are those which comprise either the sequences SEQ ID NO: 33 to 39 or the sequences SEQ ID NO:65, 66, 34-39, it being possible for SEQ ID NO:35 to be replaced by SEQ ID NO:67 and/or SEQ ID NO:36 to be replaced by SEQ ID NO:68. These haemocyanin polypeptides are particularly preferably haemocyanin 1 or 2 from *Haliotis tuberculata*.

Haemocyanin 1 from *Haliotis tuberculata*, which has an apparent molecular weight of 370 kDa in SDS-PAGE under reducing conditions, is particularly preferred.

Haemocyanin 2 from *Haliotis tuberculata*, which has an apparent molecular weight of 370 kDa in SDS-PAGE under reducing conditions, is furthermore particularly preferred. The haemocyanins are obtainable from whole haemocyanin from *Haliotis tuberculata* by the selective dissociation process described in the examples.

Haemocyanin polypeptides which are furthermore preferred are those which comprise each of the sequences SEQ ID NO: 40 to 43 or the sequences SEQ ID NO:40 to 43 and SEQ ID NO:71 to 73, it being possible in each case for the sequence with SEQ ID NO:40 to be replaced by SEQ ID NO:66 and/or SEQ ID NO:43 to be replaced by SEQ ID NO:70. Haemocyanin polypeptides which are also preferred are those which comprise either each of the sequences SEQ ID NO: 44 to 48 or the sequences SEQ ID NO:44 to 46, 77, 78, 47, 48, it being possible in each case for the sequence with SEQ ID NO:44 to be replaced by SEQ ID NO:74, SEQ ID NO:45 to be replaced by SEQ ID NO:75, SEQ ID NO:46 to be replaced by SEQ ID NO:76 and/or SEQ ID NO:47 to be replaced by SEQ ID NO:79.

These haemocyanin polypeptides are particularly preferably complete haemocyanin 1 (KLH1) or 2 (KLH2) from *Megathura crenulata*.

Non-glycosylated and glycosylated haemocyanin polypeptide obtainable by expression in host cells which are capable or incapable of glycosylation is furthermore provided. Depending on the envisaged use of the haemocyanin polypeptide, the glycosylation pattern of yeast, in particular methylotrophic yeast, of plant cells or of COS or HeLa cells can be preferred.

The invention furthermore relates to pharmaceutical compositions which comprise the nucleic acid molecules according to the invention and physiologically tolerated additives known in the prior art. The pharmaceutical compositions are preferably employed for non-specific immunostimulation in the form of a gene therapy, haemocyanin polypeptides being expressed after transformation with a suitable vector and serving to antigenize the tissue.

In particular, the invention provides the use of a nucleic acid molecule according to the invention which is bonded to an antigen-coding DNA sequence for specific immunization against this antigen. Without being bound to this theory, the immunization here is based on non-specific stimulation of the immune system by haemocyanin polypeptide epitopes and more extensive specific immunization by recognition of antigen epitopes by the immune system.

Such an immunization is particularly valuable in respect of pathogen antigens, and especially in respect of tumour antigens. The usability of the pharmaceutical composition according to the invention for treatment of tumour diseases also results from the cross-reactivity of the haemocyanin-specific antibodies with carbohydrate residues, which occur on the surface of tumours, such as e.g. the Thomsen-Friedenreich antigen, which occurs in the majority of human tumours, such as epithelial carcinomas, ovarian carcinoma, colorectal carcinoma, mammary carcinoma, bronchial carcinoma and bladder carcinoma.

The pharmaceutical compositions according to the invention can furthermore be employed for treatment of parasitic diseases, such as schistosomiasis, and for prevention of cocaine abuse.

Pharmaceutical compositions which comprise a haemocyanin polypeptide according to the invention in combination with one or more physiologically tolerated additives are provided as a further embodiment of the present invention. As already mentioned above, such a haemocyanin polypeptide can consist of a complete haemocyanin subunit, of one or more domains and of one or more fragments of such domains, provided that these fragments still have the immunological properties of at least one domain of a haemocyanin. Such a pharmaceutical composition is suitable e.g. as an antiparasitic composition, antivirus composition or antitumour composition due to either the non-specific immunostimulation, which is to be attributed solely to the haemocyanin, or due to the specific immune reaction to antigens associated with the haemocyanin. It can thus be employed e.g. for treatment of schistosomiasis, epithelial carcinomas, ovarian carcinoma, colorectal carcinoma, mammary carcinoma, bronchial carcinoma and bladder carcinomas, but is also suitable for treatment of high blood pressure. The

treatment of high blood pressure is achieved by carrying out an immunization with the aid of haemocyanin- β -adrenergic receptor peptide constructs and/or fusion proteins.

In another embodiment, the pharmaceutical compositions according to the invention are used as vaccines. They can thus make a valuable contribution to the prophylaxis of diseases caused by known pathogens. This applies in particular to pharmaceutical compositions in which a haemocyanin polypeptide is coupled to a virus, virus constituent, killed bacteria, bacteria constituents, in particular surface proteins from virus or bacteria envelopes, DNA, DNA constituents, inorganic or organic molecules, e.g. carbohydrates, peptides and/or glycoproteins.

According to another preferred embodiment, the pharmaceutical composition according to the invention is used for prevention of cocaine abuse.

Liposomes are particularly suitable for administration both of the nucleic acid molecules according to the invention and of the haemocyanin polypeptides. The present invention accordingly relates to liposomes which comprise a nucleic acid molecule according to the invention, a construct according to the invention or a haemocyanin polypeptide according to the invention.

Various methods for the preparation of liposomes which can be used for pharmaceutical purposes are known to the skilled person. The selectivity of the liposomes comprising the nucleic acid molecules or haemocyanin polypeptides according to the invention can be increased by the additional incorporation into the liposome of cell recognition molecules, which bind selectively to target cells. Receptor ligands which bind to receptors of the target cells or, especially in the case of tumours, antibodies directed against surface antigens of the particular target cells envisaged are particularly suitable for this.

The haemocyanin polypeptides according to the invention are furthermore envisaged as carrier molecules for medicaments, such as e.g. cytostatics. The increase in the

molecular weight prolongs the physiological half-life of the medicaments considerably since the loss due to ultrafiltration in the kidneys is significantly reduced.

The vaccines are formulated by methods known to the skilled person; in some embodiments the additional use of adjuvants, such as e.g. Freund's adjuvant or polysaccharides, is envisaged.

The invention furthermore provides antibodies which react specifically with the haemocyanin polypeptide according to the invention and are obtainable by immunization of a test animal with a haemocyanin polypeptide. Polyclonal antibodies can be obtained by immunization, for example, of rabbits and subsequent isolation of antisera. Monoclonal antibodies can be obtained by standard methods by immunization of e.g. mice, isolation and immortalization of the spleen cells and cloning of the hybridomas which produce antibodies specific for haemocyanin.

A screening method for identification of tumour-specific DNA in a cell is furthermore provided, this comprising the steps:

- a) bringing cell DNA and/or cell protein into contact with a probe comprising the nucleic acid molecule according to the invention and/or the antibody according to the invention and
- b) detecting the specific binding.

The tumour to be detected is preferably a bladder carcinoma, epithelial carcinoma, ovarian carcinoma, mammary carcinoma, bronchial carcinoma or colorectal carcinoma.

It is intended to illustrate the invention with the following figures and examples, but not to limit this in any way. Further embodiments, which are also included, are accessible to the skilled person on the basis of the description and the examples.

Fig. 1 shows the characterization and purification of *Haliotis tuberculata* haemocyanin (HtH):

- (a) Electron microscopy of negatively stained whole HtH, which has been purified by ultracentrifugation of cell-free haemolymph;
- (b) SDS polyacrylamide gel electrophoresis (7.5% polyacrylamide) of HtH1 compared with KLH (MW 370 kDa);
- (c) Native polyacrylamide gel electrophoresis (5% polyacrylamide) of the HtH subunit preparation, the anode being at the lower edge;
- (d) Crossed immunoelectrophoresis of the two HtH subunits using anti-HtH antibodies from the rabbit;
- (e) Electron microscopy of the remaining HtH1 didecamers (white arrows) after selective dissociation of HtH2 (black arrows);
- (f) Elution profile of the gel filtration chromatography (Biogel A15m) in the presence of ammonium molybdate/polyethylene glycol solution (pH 5.9) after selective dissociation of HtH2 into its subunit and subsequent concentration of HtH1 by ultracentrifugation;
- (g) Native polyacrylamide gel electrophoresis (6.5% polyacrylamide) of HtH1 and HtH2 subunits purified by gel chromatography compared with the starting material;
- (h,i) Crossed immunoelectrophoresis of chromatographically purified HtH subunits; and
- (j,m) Crossed immunoelectrophoresis of the purified HtH subunits using anti-KLH antibodies from the rabbit which are specific for KLH1 and KLH2.

Fig. 2 shows the analysis of the subunit organization of HtH1, anti-HtH1 antibodies from the rabbit having been used for the immunoelectrophoresis and the anode being on the left-hand side;

- (a) Crossed immunoelectrophoresis after limited proteolysis of HtH1 with the aid of elastase;
- (b) SDS polyacrylamide gel electrophoresis (7.5% polyacrylamide) of the elastase-cleaved HtH1 subunit;
- (c,d,g-j,l,n,p) Crossed immunoelectrophoresis of the elastase cleavage products of the HtH1 subunit;
- (e) Crossed immunoelectrophoresis after limited proteolysis of HtH1 with the aid of V8 protease;

- (f) SDS polyacrylamide gel electrophoresis (7.5% polyacrylamide) of the V8 protease-cleaved HtH1 subunit;
- (k,m,o) Crossed immunoelectrophoresis after limited proteolysis of HtH1 with the aid of the three stated proteases.

Fig. 3 shows the separation of proteolytic cleavage products of the subunit HtH1 with the aid of HPLC.

Fig. 4 shows the genomic sequence of the HtH1 gene.

Fig. 5 shows the primary structure deduced for the HtH1 protein.

Fig. 6 shows the genomic sequence of the HtH2 gene.

Fig. 7 shows the primary structure deduced for the HtH2 protein.

Fig. 8 shows the genomic sequence of the KLH1 gene.

Fig. 9 shows the primary structure deduced for the KLH1 protein.

Fig. 10 shows the genomic sequence of the KLH2 gene.

Fig. 11 shows the primary structure deduced for the KLH2 protein.

EXAMPLES

Material and methods

1. Preparation of the haemolymph and isolation of haemocyanin

Individuals of the European abalone *Haliotis tuberculata* from the French Atlantic coast region were provided by S.M.E.L (Blainville sur Mer, France) and Biosyn (Fellbach, Germany). The animals were kept in a 300 l sea-water aquarium at 17°C and fed with

brown algae. For removal of the haemolymph, the abalones were placed on ice in a closed plastic bag. After one hour, large volumes of haemolymph had been secreted through their skin. It emerged that the haemocyanin obtained by this process is identical to the haemocyanin which could be collected by cutting a hollow in the foot of cooled-down sea snails using a scalpel blade. The blood cells were separated from the haemolymph by centrifugation at 800 g for 30 min at 4°C. The whole haemocyanin was then immediately sedimented by preparative ultracentrifugation at 30,000 g for 4 hours at 4°C. The supernatant was discarded and the blue haemocyanin pellet was suspended overnight in "stabilization buffer" (0.05 M Tris, 5 mM CaCl₂, 5 mM MgCl₂, 0.15 M NaCl, 1 mM PMSF, pH 7.4) and stored at 4°C.

Using the process described by Harris et al., 1995, *supra*, intact HtH1 was obtained from the whole HtH by selective dissociation of HtH2 in ammonium molybdate/polyethylene glycol (1%/0.2%) solution, pH 5.9 and subsequent ultracentrifugation. The partly purified HtH1 pellet formed was dissolved and purified to homogeneity by gel filtration on a Biogel A15m device. The last step resulted in small amounts of purified HtH2. Native HtH1 and HtH2 was dissociated quantitatively into the subunits by dialysis against "dissociation buffer" (0.13 M glycine/NaOH, pH 9.6) at 4°C overnight; the presence of EDTA was not necessary. 1 mM PMSF was added at each stage of the purification to inhibit proteolysis.

2. Electron microscopy

Conventional "negative staining" was carried out by the individual drop method (Harris and Horne in Harris, J.R. (editors) *Electron microscopy in biology*, (1991), IRL Press Oxford, p. 203-228). Carbon carrier films were initially subjected to glow discharge for 20 seconds to render them hydrophilic and adsorptive for the protein. The protein samples are allowed to adsorb on to the carbon films for 60 seconds. The buffer salts are then removed by sequential washing with four successive 20 µl drops of water. Finally, the gratings are negatively stained with a 20 µl drop of 5% aqueous ammonium molybdate containing 1% trehalose (pH 7.0) and left to dry at room temperature. A Zeiss EM 900 transmission electron microscope is used for the electron microscopy analysis.

3. Polyacrylamide gel electrophoresis and immunolectrophoresis

SDS polyacrylamide gel electrophoresis (SDS-PAGE) was carried out by the method of Laemmli (Nature 227 (1970), 670-685). An alkaline system according to Markl et al. (1979) J. Comp. Physiol. 133 B, 167-175 with a 0.33 M Tris/borate, pH 9.6 as the gel buffer and 0.065 M Tris/borate, pH 9.6 as the electrode buffer was used for the native PAGE. Crossed and "crossed-line" immunolectrophoresis (IE) were carried out in accordance with Weeke (Scand. J. Immunol. 2 (1973), Suppl. 1, 47-56) or Kroll (Scand. J. Immunol. 2, Suppl. 1 (1973), 79-81). Rabbit antibodies against dissociated whole HtH and purified HtH1 were produced by Charles River Deutschland (Kisslegg, Germany). The immunization process was carried out in accordance with Markl and Winter (J. Comp. Physiol. 159B (1989), 139-151).

4. Limited proteolysis and isolation of the fragments

The limited proteolysis was carried out at 37°C in 0.13 M glycine/NaOH, pH 9.6 by addition of one of the following enzymes (Sigma, Deisenhofen, Germany), which were dissolved in 0.1 M NH₄HCO₃, pH 8.0: *Staphylococcus aureus* V8 protease type XVII (8400), papain type II from papaya milk (P-3125), bovine pancreas elastase type IV (E-0258), chymotrypsin and trypsin. The haemocyanin concentration was between 1 and 10 mg/ml. The final concentration of the enzyme was 2% (weight/weight). The proteolysis was ended after 5 hours by freezing to -20°C. The HPLC process was carried out on a device from Applied Biosystems (BAI, Bensheim, Germany) equipped with a model 1000S Diode Array detector. The proteolytic fragments were introduced on to a small Mono-Q anion exchanger column (Pharmacia, Freiburg, Germany), which had been equilibrated with 0.02 M Tris/HCl, pH 8.0, and were eluted with a linear sodium chloride gradient (0.0 M – 0.5 M NaCl) in the same buffer at a flow rate of 1 ml/min. Alternatively, the proteolytic fragments were isolated by cutting out the bands from native PAGE gels (Markl et al., 1979) J. Comp. Physiol. 133 B, 167-175, after they had first been inversely stained with the Roti-White system (Roth, Karlsruhe, Germany) in accordance with Fernandez-Patron et al. (1995) Anal. Biochem. 224, 203-211. For subsequent cleavage with a second enzyme, the fragments isolated were first dialysed overnight against 0.13 M glycine/NaOH, pH 9.6 to remove NaCl.

5. Amino acid sequence analysis

The proteins obtained by the HPLC process were denatured in SDS-containing sample buffer and separated by SDS-PAGE (Laemmli, 1970, *supra*; 7.5 % polyacrylamide). To prevent blocking of the NH₂ terminus, 0.6% (weight/weight) thioglycollic acid was added to the cathode buffer (Walsh et al., *Biochemistry* 27 (1988), 6867-6876). The protein bands were transferred by electro-transfer to ProBlot membranes (Applied Biosystems, Germany) in a vertical blotting chamber (25 mM borate buffer, pH 8.8, containing 2 mM EDTA; 10 min/100 mA, 15 min/200 mA, 12 h/300 mA). Detection of the individual polypeptides on the membranes was carried out with Ponceau S stain. The polypeptide bands of interest were cut out and sequenced in a 477A protein sequencing device from Applied Biosystems. The amounts of polypeptides applied to the sequencing device were in the lower pmol range.

6. cDNA cloning and sequence analysis

A lambda-cDNA expression library was established from poly(A⁺)-RNA from *Haliotis* mantle tissue using the vector Lambda ZAP Express[®] in accordance with the manufacturer's instructions (Stratagene, Heidelberg, Germany). The clones were isolated using HtH-specific rabbit antibodies. The nucleotide sequencing was carried out on both strands using the Taq Dye deoxy Terminator[®] system. The sequences were arranged with the software CLUSTAL W (1.7)[®] and TREEVIEW[®] (Thompson et al., *Nucl. Acids Res.* 22 (1994), 4673-4680).

Example 1:

Isolation of HtH and separation of two different types (HtH1 and HtH2)

The haemolymph was obtained from adult abalones. The blood cells were removed by centrifugation and the haemocyanin was then sedimented by ultracentrifugation. The blue haemocyanin pellet was dissolved again in "stabilization buffer" (pH 7.4) and examined by electron microscopy (figure 1a). It comprised mainly typical di-decamers, accompanied by a small content of decamers and tridecamers. Denaturing in 2% SDS in

the presence of reducing substances and subsequent SDS-PAGE separation resulted in a single band, which corresponded to the polypeptide with an apparent molecular weight of 370 kDa, which is only slightly below the apparent subunit weight of KLH (figure 1b). Complete dissociation of the oligomers and of the di-decamers into the native polypeptides (subunits) was achieved by overnight dialysis of HtH against "dissociation buffer" (pH 9.6). The native PAGE method, which was used on these samples, showed a main and a secondary component (figure 1c). Crossed immunoelectrophoresis (crossed IE) using polyclonal rabbit antibodies generated against purified whole HtH showed two components which are immunologically different but show the classical reaction of being partly immunologically identical (figure 1d). Their preparative isolation (figure 1e-i) showed that they are subunits of two different HtH types, called HtH1 and HtH2, and the patterns of the native PAGE and crossed IE methods could be assigned to each individually (figure 1c, d).

The separation of HtH1 and HtH2 was carried out by the method of selective dissociation according to Harris et al., 1995, *supra*. In ammonium molybdate/polyethylene glycol, HtH1 in the oligomer state (di-decamer) was completely stable, while HtH2 dissociated completely into the subunits (figure 1e). This allowed quantitative sedimentation of HtH1 in an ultracentrifuge, while the majority of the HtH2 remained in the supernatant. Large amounts of HtH1 were purified to homogeneity from the redissolved pellet by gel filtration chromatography, which also resulted in small amounts of pure HtH2 (figure 1f). The fractions were investigated by native PAGE (figure 1g) and crossed IE (figure 1h, i). The process of selective dissociation of HtH2 removed all the tri-decamer from the samples, which suggests that the latter are built up from HtH2, but not from HtH1 (figure 1e). The selective dissociation behaviour of HtH2 and also the ability to form aggregates which are larger than *in vivo* di-decamers correspond to the properties of KLH2. Conversely, the stability of HtH1 under these conditions and its inability to assemble into aggregates larger than di-decamers resemble the behaviour of KLH1. This feature of being related is demonstrated further by the reaction of anti-KLH1 and anti-KLH2 antibodies against the two HtH types (figure 1j-m).

Example 2:**Analysis of the organization of the HtH1 subunit**

The eight functional units (FUs, often called "functional domains") which form a mollusc haemocyanin subunit differ in primary structure and show no immunological cross-reactivity, as emerged from crossed IE. In the case of the purified HtH1 subunit (Figure 1g, h), small concentrations of five different proteases (elastase, V8 protease, papain, trypsin and chymotrypsin) which had cleaved the peptide bonds between adjacent FUs of KLH1 and KLH2 were used (Gebauer et al., 1994, *supra*, Söhngen et al., 1997, *supra*). The cleavage products were investigated by crossed IE and SDS-PAGE (Fig. 2). Elastase treatment produces eight individual FUs, deduced from the number of different immunoprecipitation peaks in the crossed IE (Fig. 2a) and with the apparent molecular weight of approx. 50 kDa of the main portion of the cleavage products in SDS-PAGE (Fig. 2b). A further precipitation peak was recognized as FU dimer, which was formed by incomplete cleavage of the segment ab (Fig. 2a). By an HPLC process with a Mono-Q column (Fig. 3a), two of the elastase cleavage products were obtained in a sufficient purity to allow their clear assignment to two of the eight precipitation peaks (Fig. 2c, d) by "crossed-line IE". The other four proteases had different cleavage patterns, which comprised mixtures of individual FUs and larger fragments containing two, three or more FUs (e.g. Fig. 2e, f). Many of them were concentrated to a sufficient amount by the HPLC process (Fig. 3b-e) to allow their identification in their corresponding SDS-PAGE and crossed IE patterns. A number of these components were sequenced N-terminally by blot transfer of SDS gels on ProBlot® membranes (Table 1). The results were compared with the N-terminal sequences which had been obtained from the apparently orthologous protein in *Megathura crenulata*, KLH1 (Table I), the complete FU arrangement of which is available (Söhngen et al., 1997, *supra*; cf. Fig. 5b). The result of the entire batch led to the determination of the complete FU arrangement within the HtH1 subunit (Fig. 2a).

In particular, cleavage of the HtH1 subunit (1-abcdefgh) with V8 protease resulted in four precipitation peaks in the crossed IE (Fig. 2e). The SDS-PAGE showed five different fragments (Fig. 2f): 220 kDa (5 FUs), 185 kDa (4 FUs), 100 kDa (2 FUs), 55 kDa (1 FU) and 46 kDa (1 FU). The 100 kDa fragment was isolated by the HPLC method (Fig. 3b)

and identified by N-terminal sequencing as 1-ab, since the sequence was identical to that of the intact subunit (Table I). In the "crossed-line" IE process, 1-ab fused with three precipitation peaks of the elastase cleavage pattern. On the basis of the evaluation, they represent fragments 1-ab, 1-a and 1-b (Fig. 2g). However, it remained unclear which peak represents 1-a and which 1-b. In a second step, the 1-ab purified by HPLC was cleaved by elastase into its component FUs, from which one could be eluted by the native PAGE gel strip method and was assigned to the elastase pattern by the "crossed-line" IE method (Fig. 2h) and sequenced N-terminally. This component had the same N-terminal sequence as the whole subunit and was therefore identical to 1-a. The second FU of the 100 kDa fragment is thus 1-b (Fig. 2a; Table I). HPLC-purified 1-c and 1-h were also obtained (Fig. 3b), identified by N-terminal sequence similarities with the corresponding FUs in KLH1 (Table I) and assigned by the "crossed-line" IE method to their corresponding precipitation peaks in the elastase pattern (Fig. 2i, j). 1-a, 1-b, 1-c and 1-h were furthermore identified (Fig. 2a). Using papain for subunit cleavage, five different peaks were obtained in the crossed IE method (Fig. 2k). A 100 kDa fragment (2 FUs) was purified from such a sample by the HPLC method (Fig. 3c), and, according to the "crossed-line" IE method, contained the FU 1-h already identified and one of the four FUs still not identified and therefore must be 1-gh (Fig. 2k, 3c). In fact, this fragment had an N-terminal sequence which showed similarities with KLH1-g (Table I). For further confirmation, the HPLC-purified fragment 1-gh was cleaved into its constituent FUs with elastase, from which 1-g was purified and identified by N-terminal sequencing. It was assigned to its peak in the elastase cleavage pattern by the "crossed-line" IE method (Fig. 2l).

The 220 kDa fragment from the V8 protease cleavage (Fig. 2e, f) was purified by HPLC (Fig. 3b) and in the "crossed-line" IE method fused with 1-h, 1-g and three peaks of the elastase cleavage pattern which have not yet been identified. The 185 kDa fragment was furthermore obtained in a sufficient purity (Fig. 2e, f; 3b), and it was shown that it comprised the same components with the exception of 1-h. This suggested that the 22 kDa and the 185 kDa fragment are 1-defgh and 1-defg respectively. In fact, the N-terminal sequence was practically identical and furthermore showed similarity with KLH1-d (Table I). Cleavage of the HtH1 subunit with trypsin resulted in a large number of components in the molecular weight range of one or two FUs (Fig. 2m). Several of the

components were concentrated in HPLC fractions (Fig. 3d). A 100 kDa fragment proved to be particularly useful since it had the same N-terminal sequence as the fragment 1-defg from the v8 protease cleavage (Table I); the 100 kDa fragment should therefore be 1-de. In the "crossed-line" IE method, this component fused with two of the three FU peaks of the elastase cleavage pattern not yet identified (Fig. 2n), which should therefore be 1-d and 1-e, and thus left a single possibility for 1-f. The "crossed-line" IE method also showed that FU 1-f was furthermore present in the 1-de fraction (Fig. 2n). The identification of 1-f was confirmed by cleavage of the subunit with chymotrypsin (Fig. 2o) and a subsequent HPLC process (Fig. 3e). This cleavage gave, *inter alia*, a 95 kDa fragment (2 FUs) which fused with 1-g and a second peak (Fig. 2p) in the "crossed-line" IE method and could therefore be either 1-gh (which could be ruled out since 1-h had already been identified) or 1-fg (which seems appropriate on the basis of the further peak in question, which was identical to the remaining candidate). In fact, this fragment showed a new N-terminal sequence which is similar to KLH1-f in a certain manner. The last problem was now to assign the two remaining FU peaks to 1-d and 1-e. This was achieved using HPLC-isolated FUs from samples in which the subunit had been cleaved with elastase. (Fig. 2c, d; 3a). The more acidic component in the crossed IE method was deduced as 1-d from its N-terminal sequence, which is identical to that of 1-defgh (Fig. 2c, Table I), while the more basic component of the 1-d/1-g pair had a new N-terminal sequence (Table I) and therefore had to be 1-e (Fig. 2a). The structure of the functional units of subunit HtH1 was thus clarified.

Example 3:

Comparison of the molecular weights and N-terminal sequences of the biochemically isolated functional units (FUs) from HtH1 and KLH1. The various FUs, each with an intact binuclear copper-binding site, were liberated from their larger unit as globular segments by limited proteolysis; cf. the section "Isolation and analysis of the units from HtH1". The KLH1 data were obtained from Söhngen et al., *supra*. The assignment as an actual unit was done on the basis of the molecular weight and the immunological properties (cf. Fig. 2). The unusually low molecular weight of isolated HtH1-d could mean that a large peptide was split off C-terminally.

TABLE 1

Functional unit	Weight (kDa)	N-terminal sequence
HtH1-a	53	DNVV RKDVS HLTDDEVQ
KLH1-a	50	ENL VRKDVER L
HtH1-b	48	?
KLH1-b	45	?
HtH1-c	46	FEDEKHSLR I RKNVDSLTPEENTNERLR
KLH1-c	45	KVPRSRL I RKNVDRLTPSE
HtH1-d	40	VEEVGTGASH I RKNLNDLNTGEM
KLH1-d	50	EVTSAN R IRKNIENLS
HtH1-e	49	ILDHDHEEEIL V RKNIIDLSP
KLH1-e	50	?
HtH1-f	50	KLNSRKHTPNR V RHELSSLSSRDIASLKA
KLH1-f	45	HHLSXNK V RHDLSTL
HtH1-g	45	DHQSGSIAGSG V RKDVTNLTKAETDNLRE
KLH1-g	45	SSMAGHF V RKDINTLTP
HtH1-h	55	DEHHDDRLLADVL I RKEVDFLSLQEANAIKD
KLH1-h	60	HEDHHEDIL V RKNIHSL

Example 4:**Cloning of haemocyanin cDNA**

1. For cloning the cDNA of haemocyanin, mRNA was isolated from the mantle tissue of the particular mollusc. The first cDNA strand was obtained by reverse transcription with Oligo(dT) as a primer. The second strand was obtained conventional synthesis with random primers. The cDNA obtained in this way was cloned in a lambda expression vector to form a cDNA expression library. Using an anti-haemocyanin antibody, the library was searched under suitable conditions, positive clones being obtained. These positive clones were isolated, sequenced and characterized.

2. A cDNA probe was prepared from the N-terminal region of a positive clone obtained, and the cDNA library was searched with this. The positive clones obtained were in turn isolated, sequenced and characterized.
3. To obtain sequences arranged still further to 5', another expression library was established from cDNA, this being obtained with the aid of a combination of haemocyanin-specific and "random" primers. This cDNA library was searched with cDNA probes which correspond to the "N-terminal" regions of the positive clones obtained under (2.). The positive clones obtained were isolated, sequenced and characterized.

Example 5:

Cloning of haemocyanin genes

Genomic DNA was isolated by standard methods. The PCR reaction was carried out with the aid of haemocyanin-specific primers in order to amplify the gene sections of the haemocyanins of interest. The amplification products obtained were cloned in a suitable vector (for example pGem T or pGem T easy (Promega, Mannheim) sequenced and characterized.

Example 6:

Recombinant expression of haemocyanin

A PCR reaction was carried out with a cDNA clone which contains the coding sequence for HtH-1d in order to amplify specifically the coding sequence of the domain 1d.

Synthetically prepared oligonucleotides were used as primers.

Primer 1 (upstream) comprises six nucleotides of the end of the domain HtH-1c, an *Sac*I cleavage site and 12 nucleotides of the end of the domain HtH-1d.

Primer 2 (downstream) comprises six nucleotides of the start of the domain HtH-1e, an *Sal*I cleavage site and an HtH1-d-specific sequence.

PCR conditions: 2 min 95°C
30 sec 95°C
30 sec 55°C
1 min 72°C
35 cycles
10 min 72°C

The amplification product was cloned in the pGEM T easy PCR cloning vector (Promega) in XL-1 Blue (Stratagene). After isolation of the recombinant plasmid and restriction with *SacI* and *SalI*, the cDNA of domain 1d could be isolated. The expression vector pQE30 (Qiagen) was also restricted with the corresponding enzymes.

The ligation was then carried out between the HtH-1d-cDNA (restricted with *SacI* and *SalI*) and pQE (restricted with *SacI* and *SalI*). Directed cloning of the cDNA which codes for HtH-1d in an expression vector is thus possible. The expression of HtH1-d in pQE in XL-1 Blue is carried out in accordance with the manufacturer's instructions. The expression of further HtH1, HtH2 or KLH1 or KLH2 domains can be carried out analogously.